Estimation of safety performance functions for urban intersections using various functional forms of the negative binomial regression model and a generalized Poisson regression model

Muhammad Wisal Khattak*
UGent, Department of Civil Engineering
Technologiepark 60, 9052 Zwijnaarde, Belgium
Tel(Admin): +32 (0) 92 64 54 89 Fax: +32 (0) 92 64 58 37
Email: Muhammad.Khattak@UGent.be
&
UHasselt, Transportation Research Institute (IMOB),
Agoralaan, 3590 Diepenbeek, Belgium
Tel (Admin): +32 (0) 11 26 91 11 Fax: +32 (0) 11 26 91 99
Email: muhammadwisal.khattak@uhasselt.be

Ali Pirdavani
UHasselt, Faculty of Engineering Technology
Agoralaan, 3590 Diepenbeek, Belgium
&
UHasselt, Transportation Research Institute (IMOB)
Agoralaan, 3590 Diepenbeek, Belgium
Tel: +32 (0) 11 29 21 83 Fax: +32 (0) 11 26 91 99
Email: ali.pirdavani@uhasselt.be

Pieter De Winne
UGent, Department of Civil Engineering
Technologiepark 60, 9052 Zwijnaarde, Belgium
Tel: +32 (0) 92 64 54 89 Fax: +32 (0) 92 64 58 37
Email: P.DeWinne@UGent.be

Tom Brijs
UHasselt, Transportation Research Institute (IMOB)
Agoralaan, 3590 Diepenbeek, Belgium
Tel: +32 (0) 11 26 91 55 Fax: +32 (0) 11 26 91 99
Email: tom.brijs@uhasselt.be

Hans De Backer
UGent, Department of Civil Engineering
Technologiepark 60, 9052 Zwijnaarde, Belgium
Tel: +32 (0) 92 64 54 34 Fax: +32 (0) 92 64 58 37
Email: Hans.DeBacker@UGent.be

*Corresponding author

Highlights

- Multiple safety performance functions (SPFs) by crash severity are developed for urban intersections
- Various functional forms of the negative binomial (NB) regression and a generalized Poisson (GP) regression model are applied to develop the SPFs
- All the NB models and a GP model show promising results when estimating the SPFs
- On the basis of goodness of fit and predictive performance measures, the developed models are compared to choose a better model
- The performance of the NB-P model is better than the competing models for signalized intersections while the GP model outperforms other models for unsignalized intersections

Estimation of safety performance functions for urban intersections using various functional forms of the negative binomial regression model and a generalized Poisson regression model

ABSTRACT

Intersections are established dangerous entities of a highway system due to the challenging and unsafe roadway environment they are characterized with for drivers and other road users. In efforts to improve safety, an enormous interest has been shown in developing statistical models for intersection crash prediction and explanation. The advantage of statistical models is that they unveil important relationships between the intersection characteristics and intersection related crashes. Accurate estimates of crash frequency and identification of crash contributing factors guide safe design and help us implement policy interventions aiming for safety improvement. In this regard, the selection of the most adequate form of crash prediction model is of great importance for the accurate estimation of crash frequency and the correct identification of contributing factors. Using a six-year crash data, road infrastructure and geometric design data, and traffic flow data of urban intersections, we applied three different functional forms of negative binomial models (NB-1, NB-2, NB-P) and a generalized Poisson (GP) model to develop safety performance functions by crash severity for signalized and unsignalized intersections. This paper presents the relationships found between the explanatory variables and the expected crash frequency and reports the comparison of different models for total, injury & fatal, and property damage only crashes to obtain those with the maximum estimation accuracy for each severity level. The comparison of models was based on both the goodness of fit and the prediction performance measures.

The fitted models showed that the traffic flow and several variables related to road infrastructure and geometric design have a significant influence on the intersection crash frequency. Further, the goodness of fit and the prediction performance measures revealed that the NB-P model outperformed other models for most of the crash severity levels in the case of signalized intersections. For the unsignalized intersections, the GP model was the best performing model. Our findings suggest a potential significant improvement in the estimation accuracy of crashes on urban intersections by applying the NB-P and GP models. Improved estimation accuracy lead to a better understanding of crash occurrence which facilitate informed decisions, effective selection and design of the countermeasures, and improve safety.

Keywords:

Urban intersections, Crash frequency, Crash severity, Negative binomial models, Safety performance functions, Geometric design

1. Introduction

Drivers encounter multiple interactions with turning and crossing vehicles, pedestrians, and cyclists at intersections. A plethora of information (e.g. the presence of road signs, street signs and name tags, traffic lights, channelization and road markings, conflicting, crossing and adjacent traffic movements, dedicated lanes for left and right turning vehicles, billboards and advert screens, etc.) at intersections produce an unsafe environment, which poses an enormous challenge for drivers to operate safely. The demand for instant decision making, complex urban design, dense and rigorous land use, congestion, heavy traffic, vulnerable road users, and many on-and-off-vehicle distractions overload the attentional resources of the driver. This in turn leads to poor judgment of the traffic situation, confusion, inadequate decision, and ultimately a crash. Hence, it is not surprising to note that intersections constitute the highest proportion of total crashes on the roads. Tay (2015) has provided some statistics from around the world to highlight this safety concern. In the past, the operational aspects of urban intersections, such as optimization of the traffic signals and/or reduction of vehicular and pedestrian traffic delays, travel time and congestion have received significant coverage in the literature (Dong et al., 2014; Roshandeh et al., 2014; Nesheli et al., 2009). However, these operational improvements do not account for the overall performance-based benefits (Roshandeh et al., 2016). The overall performance of the roadway network requires consideration of additional aspects like safety, comfort, cost, availability, accessibility, etc. In this paper, we have focused on the safety of intersections in urban areas.

The safety of intersections can be improved by understanding the factors that contribute to the occurrence of crashes and thereby, proposing appropriate countermeasures. Concerning this, an intersection safety analysis is typically suggested. One of the tools to measure the safety performance of intersections is by developing crash prediction models (CPMs). The CPMs are mathematical equations obtained through the statistical modeling of crash data and a series of explanatory variables, and are used to estimate the expected average crash frequency of roadway facilities over a specified period. They are also known as safety performance functions (SPFs) or collision prediction models (CPMs). The SPFs are applied to evaluate the safety of intersections and road segments, identify hazardous locations, assess the safety of applied solutions, and compare and prioritize the best alternative designs (AASHTO, 2010). To address safety issues, the SPFs have been developed for many years now across the globe for numerous highway facilities (Elvik et al., 2019; Abdel-Aty et al., 2016; Janstrup, 2016; Cafiso et al., 2012; Persaud et al., 2012; Vieira Gomes et al., 2012; Srinivasan and Carter, 2011; Wong et al., 2007; Greibe, 2003). Leaving aside the applicability of those models, the development of the SPFs is a critical process in which a modeler makes crucial decisions. To emphasize, Hauer and Bamfo (1997) argued, "*In the course of modeling, the modeler will make two major decisions: (a) What explanatory variables to include in the model equation; and, (b) What should be its functional form*". Factors, such as the purpose of the SPF, the availability, quality, and quantity of the data, required expertise, etc. affect those decisions.

American Association of State Highways and Transportation Officials (AASHTO) published the Highway Safety Manual (HSM), first in 2010 (AASHTO, 2010), and then in 2014 with a few supplements (AASHTO, 2014). The HSM offers the SPFs for prediction of intersection and road segment crashes on several highway facility types, e.g., rural two-lane and multilane highways, urban and suburban arterial and freeway ramp terminals (AASHTO, 2014; AASHTO, 2010). The predictive models in the HSM were developed using data from a small number of States. Because of the possible differences in the travel behavior, traffic conditions and road characteristics across different geographical regions, it has been highlighted that the crash relationships in these states may not be necessarily representative of those in the other states. Regarding this, the HSM guidelines recommend (i) the calibration of the HSM base models for applications in other jurisdictions or (ii) the estimation of new SPFs for the regions where a sufficient good quality local data is available. Several states in the US and other countries have thus developed their own SPFs. The SPFs given in the HSM for intersections estimate only total crashes that might not be an ideal approach since crashes vary by type and severity across intersections (Wang et al., 2019; Zhao et al., 2018; Wang et al., 2017). Some intersection might be crowded by fatal crashes only and others might experience injury or property damage only (PDO) crashes. Similarly, some intersections could have a higher proportion of a different particular type of crash compared with other intersections. Differences in the distribution of crash severity and/or crash type could be attributed to the variation in the geometric design and traffic characteristics between intersections. In order to consider those variations, studies estimate predictive models for intersections by crash type (Wang et al., 2019; Gates et al., 2018; Liu and Sharma, 2018; Wu et al., 2018; Dixon et al., 2015; Geedipally and Lord, 2010), and/or by severity level (Liu and Sharma, 2018; Wang et al., 2017; Wu et al., 2013; Oh et al., 2010).

Regarding the statistical methodologies, the crash prediction modeling has come a long way. In the beginning, researchers used linear regression models for the estimation of crashes and determining the relationships between crash frequency and explanatory variables (Joshua and Garber, 1990; Okamoto and Koshi, 1989). However, with new research, it was soon realized that linear regression models have certain limitations in treating the non-negative and discrete nature crash data (Lord and Mannering, 2010; Miaou and Lum, 1993). This led to the adoption of count data models in crash prediction. Naturally, the first choice of researchers was the Poisson regression model which assumes that the variance of the data is equal to the mean of the data. On the other hand, the crash data is frequently characterized by over-dispersion, that is, the variance of the crash data is greater than its mean. To overcome the over-dispersion issue, the negative binomial (NB) regression models were used (Abdel-Aty & Radwan, 2000; Miaou, 1994). With the progress in statistical methods and improved computing power, more advanced techniques have been applied recently to model the crash data. Lord and Mannering (2010), and Mannering and Bhat (2014) have provided detailed accounts of the existing trends in the crash prediction and future directions. Despite all the intricacy, the traditional NB models still enjoy great popularity due to their inherent simplicity of estimation and a relatively better performance.

Several parameterizations of the NB models are available in the literature. Nonetheless, the NB-1 and NB-2 (Cameron and Trivedi, 1986) have been commonly used to model the count data (Wang et al., 2019; Giuffrè et al., 2014; Ismail and Zamani, 2013; Hilbe, 2011; Winkelmann, 2008; Chang and Xiang, 2003; Miaou and Lord, 2003). The two models necessarily differentiate on the basis of the relationship between the variance of the data and the mean of the data. The NB-1 assumes a linear relationship between the variance and the mean, while the NB-2 assumes a quadratic relationship.

Detailed estimation procedures of the two alternative forms are given in Hardin (2018), Lord and Park (2015), and Hilbe (2011). In traffic safety, the NB-2 has been frequently used to estimate the SPFs while the NB-1 has also found a few applications. For instance, Chang and Xiang (2003) created SPFs using both the NB-1 and NB-2 models to study the relationship between crashes and congestion levels on freeways. The authors found that both models showed consistent results for the relationship between crashes and traffic volume, the number of through lanes, and median. Giuffrè et al. (2014) applied the NB-1 and NB-2 models to develop the SPFs for urban unsignalized intersections. They found that the NB-1 fits the data better than the NB-2. Wang et al. (2019) also used the NB-1 and NB-2 along with standard Poisson regression and an NB-P model for estimation of the SPFs for rural two-lane intersections.

The applications of the NB-1 and NB-2 models, however, come with a few compromises. For instance, the NB-1 and NB-2 models both restrict the variance structure in the estimation of the SPFs (Park, 2010), that is, the mean-variance relationship of the crash data is constrained to either a linear or quadratic for the NB-1 and NB-2 models, respectively. The restricted variance structure may result in the biased estimates of model parameters and ultimately the incorrect crash forecasts (Wang et al., 2019). Furthermore, both the NB-1 and NB-2 are non-nested models and an appropriate statistical test to determine a better model of the two cannot be carried out directly (Wang et al., 2019; Greene, 2008). To account for that, Greene (2008) introduced a new functional form of the NB regression called an NB-P that nests both the NB-1 and NB-2 models. The NB-P is essentially the extension of the traditional NB models to address the restricted variance structure problem. The NB-P reduces to NB-1 when P=1 and to NB-2 when P=2. Since the NB-P model parametrically nests both the NB-1 and NB-2 models, it allows analysts to test the two NB functional forms (NB-1, NB-2) against a more general alternative (NB-P) for a better model (Greene, 2008; Ismail and Zamani, 2013; Hilbe, 2011). The NB-P model has been used in a few studies dealing with count data. For example, Greene (2008) applied the NB-P along with the NB-1 and NB-2 models to the German health care data. It was found that the NB-P outperformed the other two models based on the goodness of fit measures. Ismail and Zamani (2013) used the NB-1, NB-2, and NB-P models to study the Malaysian private car own damage claim counts.

They also reported that the NB-P model was the best performing model. In traffic safety, Wang (2019) used the NB-P along with Poisson, NB-1, and NB-2 models to study the safety performance of rural two-lane intersections by crash type and intersection type. They developed traffic only models. Their findings revealed that the NB-P model performed better than the Poisson model, NB-1, and NB-2 models for most crash types and intersection types. The authors concluded that the flexible variance structure of the NB-P model significantly improves the estimation accuracy.

The literature review shows that the applications of the NB-P model, despite the obvious improvement compared to the traditional NB models, are still not common in traffic safety and crash prediction. To the authors' knowledge, no study has used the NB-P model to estimate SPFs for urban roads. Moreover, there has been no evidence that the NB-P model is used in the estimation of fully specified SPFs. Given that the applications of the NB-P model in road safety are rare, its potential to improve the estimation accuracy by offering a flexible variance structure, and the fact that it allows to statistically test the NB-1 and NB-2 against a general alternative, are motivations behind this work. Besides, the HSM recommendation of developing local SPFs for locations with enough data was another driving force. In this paper, we applied different functional forms of the NB regression model (NB-1, NB-2 and NB-P) and compare the results with the Generalized Poisson (GP) regression model, also a popular count data modeling technique, in the pursuit of obtaining the best model for the estimation of intersection SPFs in the urban areas. The GP model, discussed in section 2.4 in details, is an extension of the Generalized NB models (Ismail and Zamani, 2013). In the past, the GP models have been applied to study road crashes (Famoye et al., 2004), shipping damage incidents (Ismail and Jemain, 2007), vehicle insurance claims (Ismail and Zamani, 2013), etc. The rationale for choosing the GP model for comparison with the NB models was that it can also accommodate the over-dispersed data equally well, has relatively less applications in the SPF estimation and the fact that it is sometimes regarded as a potential competitor to the NB models for treatment of over-dispersed count data (Melliana et al., 2013). The contribution of the current study to traffic safety literature is that it applies the functional form NB-P of the NB regression, along with the NB-1, NB-2 and a GP model for the estimation of intersection SPFs in the urban areas. A unique combination of the new approach for the

SPFs estimation and the use of not only the traffic flow but also other explanatory variables adds to the novelty of this work. To the best of our knowledge, no micro-level SPFs have been developed for the urban intersections in Belgium, results of this study could potentially serve the local research community involved in traffic safety as well as the industry in planning level safety assessment of new road infrastructure projects.

2. Methodology

The count data models have been widely applied to estimate crashes at the road segments and intersections in a non-negative, discrete, and random fashion (Washington et al., 2010). Since the Poisson regression models are usually not fit for modeling the crash data due to their inability to accommodate overdispersion, three different functional form of the NB model and a GP model were applied to estimate the SPFs for urban intersections in this study.

2.1 Negative binomial model-type 2 (NB-2)

The negative binomial regression is the derivative of the standard Poisson regression. It redefines the conditional mean of the standard Poisson model (equi-dispersion; variance of the data equals its mean) and incorporates a latent heterogeneity term to account for over-dispersion in data. The expected crash frequency "μ_i" at the intersection "i" obtained by applying the NB model as in Washington et al. (2010) is given by:

$$\mu_i = exp(\beta X_i + \varepsilon_i) \quad (1)$$

where "X_i" is the vector of explanatory variables, "β" is the vector of estimable coefficients and "exp(ε_i)" is the latent heterogeneity term, also known as an error term. When the "exp(ε_i)" follows gamma distribution with mean 1 and variance $1/\sigma = k$ where "k" represents an over-dispersion parameter, a traditional NB model, called the NB-2 model, is derived.

For the interest of readers, an equation 1 according to the standard Poisson regression model would have been:

$$\mu_i = exp(\beta X_i) \quad (2)$$

Clearly, this lacks the term "exp(ε_i)" to account for over-dispersion.

The probability density function of the NB-2 model for estimation of the SPFs as in Washington et al. (2010):

$$Prob[y_i|\mu_i] = \frac{\Gamma[(\sigma) + y_i]}{\Gamma(\sigma)y_i!}\left[\frac{\sigma}{(\sigma) + \mu_i}\right]^{\sigma}\left[\frac{\mu_i}{(\sigma) + \mu_i}\right]^{y_i} \tag{3}$$

where Γ is a gamma function. The mean and the variance of the NB-2 regression model are equal to $E(y_i) = \mu_i$, and $Var(y_i) = \mu_i + k\mu_i^2 = \mu_i(1 + k\mu_i)$, respectively. When $1/\sigma = k$, the marginal distribution function of the NB-2 model can be reproduced:

$$Prob[y_i|\mu_i] = \frac{\Gamma\left[\left(\frac{1}{k}\right) + y_i\right]}{\Gamma\left(\frac{1}{k}\right)y_i!}\left[\frac{\frac{1}{k}}{\left(\frac{1}{k}\right) + \mu_i}\right]^{\frac{1}{k}}\left[\frac{\mu_i}{\left(\frac{1}{k}\right) + \mu_i}\right]^{y_i} \tag{4}$$

2.2 Negative binomial model-type 1 (NB-1)

A re-parameterization of the variance structure of the NB model by replacing $\frac{1}{k}$ in the NB-2 (equation 4) with $\frac{1}{k}\mu_i$ allows for another functional form, called the NB-1 (Wang et al., 2019; Hilbe, 2011; Greene, 2008; Cameron & Trivedi, 1986). The marginal distribution function of the NB-1 is given by:

$$Prob[y_i|\mu_i] = \frac{\Gamma\left[\left(\frac{1}{k}\mu_i\right) + y_i\right]}{\Gamma\left(\frac{1}{k}\mu_i\right)y_i!}\left[\frac{\frac{1}{k}\mu_i}{\left(\frac{1}{k}\mu_i\right) + \mu_i}\right]^{\frac{1}{k}\mu_i}\left[\frac{\mu_i}{\left(\frac{1}{k}\mu_i\right) + \mu_i}\right]^{y_i} \tag{5}$$

The mean of the NB-1 is $E(y_i) = \mu_i$ and the variance of the NB-1 is $Var(y_i) = \mu_i + k\mu_i$.

2.3 Negative binomial model-type P (NB-P)

Greene (2008) proposed a new form of the NB regression that uses the parameter "P" to represent the mean-variance relationship. It is known as the NB-P model. The NB-P model is obtained by replacing $\frac{1}{k}$ in the NB-2 model (equation 4) with $\frac{1}{k}\mu_i^{\,2-P}$. The marginal distribution function of the NB-P model is given by:

$$Prob[y_i|\mu_i] = \frac{\Gamma\left[\left(\frac{1}{k}\mu_i^{\,2-P}\right) + y_i\right]}{\Gamma\left(\frac{1}{k}\mu_i^{\,2-P}\right)y_i!}\left[\frac{\frac{1}{k}\mu_i^{\,2-P}}{\left(\frac{1}{k}\mu_i^{\,2-P}\right) + \mu_i}\right]^{\frac{1}{k}\mu_i^{\,2-P}}\left[\frac{\mu_i}{\left(\frac{1}{k}\mu_i^{\,2-P}\right) + \mu_i}\right]^{y_i} \tag{6}$$

where mean and variance of the NB-P are $E(y_i) = \mu_i$ and $Var(y_i) = \mu_i + k\mu_i^P$, respectively. "P" represents the functional parameter of the NB-P model.

All the NB models used maximum likelihood estimation (MLE) approach to estimate the parameter coefficients.

2.4 Generalized Poisson model (GP)

The generalized Poisson (GP) regression is another popular approach to model count data. As an alternative to the NB regression, the GP models have the advantage of modeling both over-dispersed and under-dispersed data. Like the NB regression, the GP model has an extra parameter, called a scale or dispersion parameter. A distinctive feature of the GP dispersion parameter is that it can take both positive and negative values for over-dispersed and under-dispersed data, respectively. The probability mass function (p.m.f.) of the GP distribution given as in Yang et al. (2009):

$$Prob[Y_i|y_i] = \frac{\theta(\theta+ky_i)^{y_i-1}\exp(-\theta-ky_i)}{y_i!}, \qquad y_i = 0,1,2,\dots, \tag{7}$$

where $\theta > 0$, and $0 \leq k < 1$. From Joe and Zhu (2005), the mean of the GP regression is $E(Y_i) = \mu = (1-k)^{-1}\theta$, and the variance of the GP regression is $Var(Y_i) = (1-k)^{-3}\theta = (1-k)^{-2}\mu = \emptyset.\mu$. The term $\emptyset = (1-k)^{-2}$ is a dispersion factor, and it is used in the GP mass function where "k" is a dispersion parameter. It can be seen that when $k = 0$, a standard Poisson model is obtained. For $k < 0$, under-dispersion is assumed while $k > 0$ represents over-dispersion. Since crash data normally exhibits over-dispersion, this study will assume $k > 0$ condition. There are other parametrizations of the GP but their applications are left for future studies.

2.5 Model structure

The literature offers several ways to model the relationships between intersection crash frequency and explanatory variables (Barbosa et al., 2014; Park and Lord, 2009; Nambuusi et al., 2008; Miaou and Lord, 2003). They are differentiated on the basis of the type of variables, the number of variables, the form that the variables take during the modeling process and the transformation applied to the variables (Oh et al., 2003). In this study, the following model structure was used to estimate the expected crash frequency "μ_i" of the intersection "i":

$$\mu_i = exp(\beta_0 + \beta_1\, ln(AADT_{major}) + \beta_2\, ln(AADT_{minor}) + \Sigma_{m=3}^{n} \beta_m X_m) \qquad (8)$$

where β_0 represents the intercept, $AADT_{major}$ is the major approach average annual daily traffic (AADT), β_1 represents the coefficient estimate of the major approach AADT, $AADT_{minor}$ represents the minor approach AADT, β_2 represents the coefficient estimate of the minor approach AADT, β_m is the vector of the coefficient estimates of explanatory variables and "X_m" denotes the vector of explanatory variables. For the NB models (NB-1, NB-2, and NB-P) and the GP model, the coefficients denoted by β_m and a dispersion parameter denoted by "k" were estimated but for the NB-P, an additional parameter "P", called a functional parameter, was also estimated.

2.6 Model comparison

For model comparison, both the likelihood-based and the predictive ability-based measures were used. The likelihood-based measures consisted of the likelihood ratio test (LRT), the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC). The LRT was used only when comparing the hierarchically nested models (Greene, 2008; Wang et al., 2019). The AIC and the BIC were used for comparing the non-nested models (Ismail and Jemain, 2007).

The predictive ability-based measures compared all developed models for predictive performance using the validation data. Those included in the study were; mean prediction bias (MPB), mean absolute deviation (MAD), and mean squared prediction error (MSPE) as in Oh et al. (2003), and % CURE deviation and a validation factor (Hauer, 2015; Wang et al., 2019).

3. Data

The data used for modelling was obtained for urban intersections of Antwerp, Belgium. A dataset consisting of crash data of six years (2010-2015), road geometric data, and traffic flow data was created for the estimation of the SPFs. An online database of the regional government called the Flanders road register was consulted for the intersection data. A total of 760 intersections were used for analysis, of which 198 were signalized and 562 were unsignalized. Around 470 were three-legged intersections and the remaining 290 were four-legged intersections. Because the skewness of intersection has been reported to have an impact on its safety (Nightingale et al., 2017; Haleem and

Abdel-Aty, 2010), it was decided to include skewness as a potential explanatory variable. The smallest angle between the two adjacent approaches of intersection, known as an intersection angle (Nightingale et al., 2017), was used as a surrogate to define the level of skewness. A 75 degrees intersection angle used by Haleem and Abdel-Aty (2010) was chosen as a threshold to define the levels of skewness. An intersection angle less than or equal 75 degrees represented skewness level 1 while an intersection angle greater than 75 degrees represented skewness level 2. A total of 217 intersections had a skewness level 1 and 543 intersections had a skewness level 2. **Table 1** provides the description of variables employed in this study for urban signalized and unsignalized intersections.

Table 1 Variables description for urban intersections of Antwerp

Variable Description	**Variable levels**
AADT on the major approach	-
AADT on the minor approach	-
Skewness	1: Intersection angle is less than/equal to 75-degrees 2: Intersection angle is greater than 75-degrees
Legs/approaches of the intersection	1: For 4 legged intersections 0: For 3 legged intersections
Existence of stop sign on the minor approach	1: Stop sign is present on at least one minor approach 0: No stop sign on the minor approaches
Existence of stop line on the minor approach	1: Stop line is present on at least one minor approach 0: No stop line on the minor approaches
Number of left turn lane on the major approach	2: At least one left turn lane exists on each direction of the major approach 1: At least one left turn lane exists on only one direction of the major approach 0: No left turn lane exists
Number of right turn lane on the major approach	2: At least one right turn lane exists on each direction of the major approach 1: At least one right turn lane exists on only one direction of the major approach 0: No right turn lane exists
Number of through lanes of the minor approach	4 or 4+: Four and more through lanes of the minor approach 1-3: One to three through lanes of the minor approach 0: No through lane of the minor approach
Left turn (LT) movements on the minor approach	2: LT movement on each minor approach 1: LT movement on only one minor approach 0: No LT movement on the minor approach

Existence of crosswalk on minor approach	2: Crosswalk on each minor approach
	1: Crosswalk on only one minor approach
	0: No crosswalk
Existence of crosswalk on major approach	2: Crosswalk on each major approach
	1: Crosswalk on only one major approach
	0: No crosswalk
Size of the intersection [a]	4: for 5*4, 5*8, 6*4, 6*6, 6*8, 8*4, 8*6, 8*8, 8*10, 10*8, 10*10
	3: for 3*2, 3*4, 3*6, 4*2, 4*4, 4*6
	2: for 2*2, 2*3, 2*4, 2*6
	1: for 1*2, 1*3, 1*4

[a] The first number is the total number of approach lanes for a minor approach, and the second number is the total number of through lanes for a major approach (as per, Abdel-Aty and Haleem 2011)

The crash data was provided by the police of Antwerp. The crash records featured the severity level of a crash, coordinates of a crash location, time and date of a crash, number of the vehicles involved and their type, maneuver of the involved vehicles at the time of the crash, data about the involved drivers, and road and pavement conditions. Only intersection and intersection-related crashes were used in the analysis. Because of the inconstancy in the definition of the influence area to classify a crash as intersection-related (Wang et al., 2008), we chose to use the HSM guidelines to differentiate the intersection and intersection-related crashes from the segment crashes. According to the HSM (AASHTO, 2014, 2010);

- An intersection crash is the one that has occurred within the physical boundaries of an intersection area
- An intersection related crash is the one that has occurred on the road segment but the presence of the intersection was the cause of that crash and it falls within its influence area

Using the above definition, 5128 intersection and intersection related crashes were identified for analysis. To account for the potential variation in the SPFs by crash severity, those crashes were divided into total crashes, injury & fatal crashes and property damage only (PDO) crashes.

The traffic data was acquired from Lantis, a mobility management company based in Antwerp. Lantis also provides its services to the Mobiliteit en Parkeren Antwerpen Ag, an office for parking and mobility services of Antwerp city. The data was received in two sets, actual counts and traffic model estimates. The actual counts were collected using either manual counting techniques or loop detectors installed at the random locations on the roads in the study network. The traffic model estimates were generated using a microsimulation traffic model called Dynamisch Model Kernstad Antwerpen (DMKA). It is important to note that the model was calibrated for the years 2010-2015, a period during which the crash data was recorded. Results from several runs of the simulation model were obtained and averaged to get a better convergence towards the actual counts. Actual counts and model generated counts were compared at locations where both were available to check for the residuals. An absolute difference of not greater than 5% between the simulation counts and actual counts was reported for the majority of locations. The outliers were discarded. The authors agreed to use a combination of actual counts and traffic model estimates to ensure as many intersections included in the SPFs estimation as possible with a maximum degree of accuracy. **Table 2** provides the descriptive statistics of crash data (by severity) and traffic data for signalized and unsignalized intersections used to develop the SPFs.

Table 2 Descriptive statistics of crash data (by severity) and traffic flow data for signalized and unsignalized intersections

Variables	Signalized Intersections				Unsignalized Intersections			
	Min.	**Max.**	**Aver.**	**Std. Dev.**	**Min.**	**Max.**	**Aver.**	**Std. Dev.**
Total Crashes	0	87	13.899	13.848	0	51	4.347	5.223
PDO Crashes	0	50	6.979	7.760	0	49	2.540	3.671
Injury & Fatal Crashes	0	39	6.919	7.224	0	25	1.806	2.557
Ln $(AADT)_{major}$	183	41915	14559	9424.8	13	30648	3511	2884.1
Ln $(AADT)_{minor}$	31	26837	5225	4905.8	9	7595	1001	815.2

4. Results

Table 3 and **Table 4** present the parameter estimates (β) of the NB-1, NB-2, NB-P, and GP models developed by crash severity (total crashes, PDO crashes, and injury & fatal crashes) for signalized and unsignalized intersections, respectively. The numbers enclosed within the parenthesis correspond to their p-values. The SPFs show that the signs of estimated parameters are similar across different models developed for the same severity level. This indicates that given the same severity level, the potential impact of explanatory variables on the expected crash frequency obtained from different models is similar. The estimated parameters, however, vary slightly across different severity levels which could be one of the reasons that imply the need to develop separate models for each crash severity level. Using a 90% confidence level as in Vieira Gomes et al. (2012) for similar data, we found that five variables were significant in case of signalized intersections and four variables in case of unsignalized intersections. The significant variables included the traffic flow, the intersection skewness, the existence of crosswalk on a minor approach, the number of through lanes on a minor approach, and the number of approaches. To our surprise, the presence of exclusive left and right turn lanes were not significant in any model. The intersection size and the crosswalk on the major approaches were other insignificant explanatory variables.

4.1 SPFs of signalized intersections

Table 3 provides the SPF estimation results for signalized intersection. It shows that there was a statistically significant increase in the crash frequency with an increase in the natural logarithm of AADTs (which necessarily indicates an increase in traffic flow) of the major and the minor approaches of intersection. The crosswalk on a minor approach was significant only when it existed on both approaches of a signalized intersection across all developed models and all severity levels. However, there was an exception in case of the NB-2 and NB-P models of total crashes, for which, in addition to a crosswalk on each minor approach, a crosswalk variable was also significant when present on only one of the minor approaches of an intersection. The estimated coefficients in the former case were approximately double than that of the later. This was not true for other crash severity levels (the PDO, and injury & fatal crashes) and model types. The intersection skewness was significant only for total crashes (all the NB models only), and injury & fatal crashes (all models). The coefficient estimates were negative in the developed models. Since the higher skewness level was a base case, the negative sign indicates that no skewness or lower skewness (i.e., intersection angle greater than 75 degrees, please see the data section for details) results in a reduced crash frequency. In other words, intersections with no or lower skewness were safer than the intersections with higher skewness. This is a straight forward result since the presence of skewness causes larger intersection areas, obstructs views and affects sight distances. An important observation from the results was that the absence of skewness causes a greater decrease in the injury & fatal crashes than the total crash frequency.

Table 3 Estimated models for urban signalized intersections

Variables	NB-1 β (p-value)	NB-2 β (p-value)	NB-P β (p-value)	GP β (p-value)
TOTAL CRASHES				
Intercept	-3.9067 (0.0000)	-4.2775 (0.0000)	-4.2761 (0.0000)	-3.7402 (0.0000)
AADT $_{Major}$	0.4058 (0.0000)	0.3623 (0.0000)	0.3621 (0.0000)	0.3934 (0.0000)
AADT $_{Minor}$	0.2450 (0.0000)	0.3002 (0.0000)	0.3003 (0.0000)	0.2425 (0.0000)
No crosswalk: 0 (Base)				
Crosswalk on one of the minor approaches: 1	0.2747 (0.3870)	0.5547 (0.0690)	0.5551 (0.0690)	0.2578 (0.3920)
Crosswalk on each of the minor approach: 2	0.8867 (0.0050)	1.2151 (0.0000)	1.2155 (0.0000)	0.8549 (0.0040)
Skewness: 1 (Base)				
Skewness: 2	-0.1572 (0.0970)	-0.2180 (0.0360)	-0.2181 (0.0360)	-0.1486 (0.1190)
Over-dispersion	4.1062	0.2977	0.2953	0.5778
P	1.000 (0.0000)	2.000 (0.0000)	2.0031 (0.0000)	
Log L[a]	-653.03	-640.65	-640.65	-651.79
AIC	1320.06	1295.31	1297.31	1317.58
BIC	1343.07	1318.33	1323.62	1340.60
PDO CRASHES				
Intercept	-4.4085 (0.0000)	-4.8088 (0.0000)	-4.8899 (0.0000)	-4.2727 (0.0000)
AADT $_{Major}$	0.3396 (0.0100)	0.2992 (0.0010)	0.3153 (0.0010)	0.3269 (0.0010)
AADT $_{Minor}$	0.2954 (0.0000)	0.3367 (0.0000)	0.3357 (0.0000)	0.2942 (0.0000)
No crosswalk: 0 (Base)				
Crosswalk on one of the minor approaches: 1	0.2377 (0.5320)	0.6326 (0.1050)	0.5671 (0.1810)	0.2382 (0.5190)
Crosswalk on each of the minor approaches: 2	0.9062 (0.0150)	1.3397 (0.0010)	1.2820 (0.0020)	0.9008 (0.0130)
Over-dispersion	2.7650	0.3840	0.6319	0.5022
P	1.000 (0.0000)	2.000 (0.0000)	1.7530 (0.0000)	
Log L	-538.89	-533.32	-532.91	-538.16
AIC	1091.79	1080.65	1081.81	1090.33
BIC	1114.80	1103.66	1108.12	1113.35
INJURY & FATAL CRASHES				
Intercept	-4.9921 (0.0000)	-5.6066 (0.0000)	-5.6210 (0.0000)	-4.9458 (0.0000)
AADT $_{Major}$	0.4963 (0.0000)	0.4797 (0.0000)	0.4835 (0.0000)	0.4952 (0.0000)
AADT $_{Minor}$	0.1917 (0.0030)	0.2586 (0.0000)	0.2563 (0.0000)	0.1879 (0.0040)
No crosswalk: 0 (Base)				
Crosswalk on one of the minor approaches: 1	0.2618 (0.4690)	0.4785 (0.2020)	0.4738 (0.2110)	0.2631 (0.0040)
Crosswalk on each of the minor approaches: 2	0.8633 (0.0150)	1.0876 (0.0030)	1.0856 (0.0040)	0.8576 (0.0140)
Skewness: 1 (Base)				
Skewness: 2	-0.2056 (0.0620)	-0.3258 (0.0070)	-0.3213 (0.0090)	-0.1984 (0.0740)
Over-dispersion	2.3591	0.3324	0.3679	0.4727
P	1.000	2.000	1.9500	

	(0.0000)	(0.0000)	(0.0000)	
Log L	-531.95	524.40	-524.39	-530.87
AIC	1077.91	1062.81	1064.77	1075.74
BIC	1100.93	1085.83	1091.08	1098.76

Notes: [a] Log L: Log Likelihood

4.2 SPFs of unsignalized intersections

Table 4 presents the coefficient estimates of the SPFs for unsignalized intersections. The traffic flows of major and minor approaches were significantly associated with crash frequency except for injury and fatal crashes where the AADT of the minor approach was found insignificant. The presence of a crosswalk on the minor approach was only significant for total crashes, and injury and fatal crashes across all developed models. Unlike signalized intersections, the crosswalk was significant when it was present on only one of the minor approaches of unsignalized intersections. The presence of crosswalk on one or both approaches was, however, significant only in case of injury and fatal crashes as can be seen in the NB-2 and NB-P models. The number of approaches/legs of an intersection was a significant predictor of total and PDO crashes at unsignalized intersections at a 90% confidence level. Intersections with three approaches/legs as a base, the positive signs of the estimated coefficients indicate higher expected crash frequency on intersections with four approaches compared to intersections with three approaches. Another statistically significant variable was the number of through lanes of the minor approaches of unsignalized intersection. A positive association was found between crash frequency and the number of through lanes of its minor approach for the total crashes, and injury & fatal crashes. While the first level of this variable was not significant, the second level, which represents four or greater number of through lanes of minor approaches was significant for total crashes. For injury & fatal crashes, all levels of the number of through lanes were significant. This means that a significant increase can be expected in total crashes, and injury & fatal crashes with an increase in the number of through lanes of the minor approach of an unsignalized intersection. It is noteworthy that this result can be generalized only to four-legged unsignalized intersections because through lanes were reported only for such facility type in this study.

Table 4 Estimated models for urban unsignalized intersections

Variables	NB-1 β (p-value)	NB-2 β (p-value)	NB-P β (p-value)	GP β (p-value)
TOTAL CRASHES				
Intercept	-1.2095 (0.0000)	-1.4860 (0.0000)	-1.4082 (0.0000)	-1.1683 (0.0000)
$AADT_{Major}$	0.1948 (0.0000)	0.2155 (0.0000)	0.2113 (0.0000)	0.1883 (0.0000)
$AADT_{Minor}$	0.1262 (0.0010)	0.1539 (0.0010)	0.1379 (0.0010)	0.1266 (0.0010)
No crosswalk: 0 (Base)				
Crosswalk on one of the minor approaches: 1	0.2668 (0.0010)	0.1709 (0.0500)	0.2485 (0.0040)	0.2728 (0.0010)
Crosswalk on each of the minor approaches: 2	0.1609 (0.1690)	0.1787 (0.1970)	0.1743 (0.1650)	0.1728 (0.1380)
No. of approaches: 3 (Base)				
No. of approaches: 4	0.3878 (0.0010)	0.2369 (0.0950)	0.3547 (0.0070)	0.3994 (0.0010)
No. of through lanes on the minor approaches: 0 (Base)				
No. of through lanes on the minor approach: 1-3	0.0330 (0.7880)	0.1080 (0.4730)	0.0578 (0.6690)	0.0260 (0.8310)
No. of through lanes on the minor approach: 4 & 4+	0.8029 (0.0000)	0.8797 (0.0120)	0.8176 (0.0010)	0.7760 (0.0000)
Over-dispersion	2.3705	0.5680	1.4209	0.4708
P	1.0000 (0.0000)	2.0000 (0.0000)	1.3598 (0.0000)	
Log L[a]	-1369.80	-1372.45	-1368.53	-1362.96
AIC	2757.61	2762.89	2757.06	2743.93
BIC	2796.61	2801.89	2800.39	2782.93
PDO CRASHES				
Intercept	-1.039 (0.0020)	-1.2663 (0.0000)	-1.2223 (0.0010)	-1.0004 (0.0030)
$AADT_{Major}$	0.1011 (0.0370)	0.0989 (0.0600)	0.1067 (0.038)	0.0987 (0.0420)
$AADT_{Minor}$	0.1619 (0.0010)	0.2040 (0.0000)	0.1842 (0.0010)	0.1583 (0.0010)
No. of approaches: 3 (Base)				
No. of approaches: 4	0.3189 (0.0000)	0.2122 (0.0280)	0.2912 (0.0030)	0.3291 (0.0000)
Over-dispersion	1.7930	0.7114	1.1867	0.4167
P	1.0000 (0.0000)	2.0000 (0.0000)	1.4507 (0.0000)	
Log L	-1149.44	-1149.70	-1148.76	-1140.98
AIC	2308.87	2309.41	2309.51	2291.96
BIC	2330.54	2331.07	2335.51	2313.63
INJURY & FATAL CRASHES				
Intercept	-3.6679 (0.0000)	-4.0729 (0.0000)	-4.0075 (0.0000)	-3.6730 (0.0000)
$AADT_{Major}$	0.4225 (0.0000)	0.4507 (0.0000)	0.4489 (0.0000)	0.4229 (0.0000)
$AADT_{Minor}$	0.0727 (0.1510)	0.0912 (0.1060)	0.0857 (0.1170)	0.0727 (0.1530)
No crosswalk: 0 (Base)				
Crosswalk on one of the minor approaches: 1	0.3155 (0.0030)	0.3582 (0.0020)	0.3439 (0.0020)	0.3136 (0.0030)
Crosswalk on each of the minor approaches: 2	0.2379 (0.1080)	0.3740 (0.0290)	0.3091 (0.0570)	0.2425 (0.1020)
No. of through lanes on the minor approaches: 0 (Base)				
No. of through lanes on the minor approaches: 1-3	0.4208	0.5060	0.4678	0.4167

	(0.0100)	(0.0080)	(0.0090)	(0.0110)
No. of through lanes on the minor approaches: 4 & 4+	1.2610 (0.0000)	1.2376 (0.0020)	1.2515 (0.0000)	1.2484 (0.0000)
Over-dispersion	1.2707	0.5680	0.9755	0.3489
P	1.0000 (0.0000)	2.0000 (0.0000)	1.4601 (0.0000)	
Log L	-931.85	-933.04	-929.64	-931.01
AIC	1881.70	1884.07	1879.28	1880.02
BIC	1920.70	1923.07	1922.61	1919.02

Notes: [a] Log L: Log Likelihood

4.3 Comparison and performance evaluation of the developed SPFs

The likelihood ratio test (LRT) was used for the comparison of either the NB-1 with the NB-P model or the NB-2 with the NB-P model since both the NB-1 and NB-2 are parametrically nested by the NB-P (Greene, 2008). The LTR was, however, not applied to compare the non-nested models, i.e., the NB-1 model against the NB-2 model, or the NB models against the GP model. Instead, the AIC and BIC were used as in Ismail and Jemain (2007).

Table 5 Likelihood ratio (NB-1 vs NB-P and NB-2 vs NB-P) for signalized and unsignalized intersections

	Signalized Intersections		**Unsignalized Intersections**	
TOTAL CRASHES				
Test/Criteria	**NB-1**	**NB-P**	**NB-1**	**NB-P**
Log L[a]	-653.028	-640.655	-1369.804	-1368.529
Likelihood ratio ($\chi 2$)		**24.75** (0.0000) [b]		2.55 (0.1104)
Test/Criteria	**NB-2**	**NB-P**	**NB-2**	**NB-P**
Log L	-640.655	-640.6552	-1372.4456	-1368.529
Likelihood ratio ($\chi 2$)		0.0002 (0.9893)		**7.83 (0.0051**)
PDO CRASHES				
Test/Criteria	**NB-1**	**NB-P**	**NB-1**	**NB-P**
Log L	-538.894	-532.906	-1149.436	-1148.756
Likelihood ratio ($\chi 2$)		**11.98** (0.0005)		1.36 (0.2436)
Test/Criteria	**NB-2**	**NB-P**	**NB-2**	**NB-P**
Log L	-533.324	-532.906	-1149.705	-1148.756
Likelihood ratio ($\chi 2$)		0.84 (0.3606)		1.90 (0.1682)
INJURY & FATAL CRASHES				
Test/Criteria	**NB-1**	**NB-P**	**NB-1**	**NB-P**
Log L	-531.954	-524.388	-931.851	-929.6407
Likelihood ratio ($\chi 2$)		**15.13** (0.0001)		**4.42** (0.0355)
Test/Criteria	**NB-2**	**NB-P**	**NB-2**	**NB-P**
Log L	524.404	-524.388	-933.036	-929.6407
Likelihood ratio ($\chi 2$)		0.03 (0.8569)		**6.79** (0.0092)

Notes: Bold values indicate statistically significant results of the LRT
[a] Log L: Log Likelihood
[b] Values in parenthesis indicate the p-value when the likelihood ratio ($\chi 2$) was computed

The LRT indicated that the NB-P model performed better than the NB-1 model for total crashes, PDO crashes, and injury & fatal crashes in case of signalized intersections (**Table 5**). The result of the LTR test was, however, inconclusive when the NB-P and NB-2 were compared and, hence, it cannot be said with certainty which of the two was a better model. Based on the other measures, i.e., log-likelihood, the AIC, and the BIC (**Table 3**), it can be seen that NB-P and NB-2 performed relatively closely but both performed better than the NB-1 models and the GP models for crash severities. The functional parameter "P" of the estimated NB-P models was statistically significant across all severity levels. The estimated value of the functional parameter "P" of the NB-P models for total crashes, and injury & fatal crashes was close to 2 while for the PDO crashes it was significantly different from either 1 or 2 (**Table 3**). Although this does not completely verify the assumption that the restricted variance structure of the NB-1 or NB-2 models may lead to biased estimates of model parameters, it does not entirely reject the possibility either, as indicated by the PDO crashes on signalized intersections and the result for the NB-1 models.

The LRT for unsignalized intersections showed that the NB-P and NB-1 models performed equally closely in case of total crashes and the PDO crashes and we cannot say that the difference in the NB-P and NB-1 estimates was significant but for injury & fatal crashes, the results were in the favor of the NB-P models. The NB-P model, on the other hand, outperformed the NB-2 model for total crashes, and injury & fatal crash but there was no significant difference in the estimates of the NB-2 and NB-P models for the PDO crashes. Based on the AIC and BIC values, the NB-1 models performed better than the NB-2 models (non-nested models comparison, **Table 4**) for total crashes and injury & fatal crashes while results for the PDO crashes were fairly close for the two traditional NB models. The AIC and BIC, however, showed better model fit for the GP models in all crash severity levels. So, it will be safe to say that the GP model outperformed all the NB models in the case of un-signalized intersections. The functional parameter "P" of variance structure was significant for the NB-P models across all severity levels and it was not close to either 1 or 2. This verifies the assumption that the restricted variance structure of the NB-1 and NB-2 models might lead to the biased estimates of model parameters for unsignalized intersection, and, hence the NB-P that takes into account the flexible variance structure would be more reliable in the accurate estimation of model parameters when there is no GP model considered.

Besides the likelihood-based criteria, predictive ability-based measures were also computed to validate the developed models and examine their predictive performance. It is important to note that randomly selected 80% data were used for the estimation of models while the remaining 20% were used for validation of the developed models. We compute the MPB, MAD, MSPE, % CURE deviation and a validation factor. According to Oh et al. (2003), smaller the absolute values of the MPB, MAD, and MSPE, better is the performance of the developed models. The % CURE deviation, which denotes the percentage of the data points falling outside the two standard deviation limits of the Cumulative Residual (CURE) (Hauer, 2015), shows a good fit when its values are small (Wang et al., 2019). Finally, a factor, that we called a validation factor, was calculated as the ratio of the total predicted crashes to the total observed crashes using the validation data. A value close to one indicated a better model (Wang et al., 2019). Wang et al., (2019) called it a calibration factor.

Table 6 Predictive performance evaluation and validation of estimated models of signalized and unsignalized intersections

Crash Severity	**Criteria**	**Signalized Intersections (198)**				**Unsignalized Intersections (562)**			
		NB-1	**NB-2**	**NB-P**	**GP**	**NB-1**	**NB-2**	**NB-P**	**GP**
TOTAL CRASHES	**MPB**	-0.233	-0.268	-0.268	-0.237	-0.035	-0.034	-0.035	-0.034
	MAD	1.083	1.082	1.082	1.088	0.509	0.510	0.507	0.509
	MSPE	2.998	2.932	2.932	3.042	0.472	0.473	0.470	0.471
	CURE Deviation (%)	26	4	4	36	0	1	0	0
	Validation Factor (VF)	1.094	1.110	1.110	1.096	0.954	0.952	0.953	0.955
PDO CRASHES	**MPB**	0.043	0.031	0.040	0.041	-0.025	-0.027	-0.026	-0.027
	MAD	0.688	0.693	0.693	0.691	0.340	0.337	0.337	0.337
	MSPE	1.100	1.106	1.097	1.111	0.419	0.420	0.419	0.418
	CURE Deviation (%)	19	5	6	21	0	0	0	1
	Validation Factor (VF)	1.033	1.024	1.031	1.032	0.946	0.943	0.944	0.944
INJURY & FATAL CRASHES	**MPB**	0.063	0.039	0.040	0.064	-0.002	0.002	-0.001	-0.002
	MAD	0.629	0.623	0.624	0.629	0.248	0.246	0.247	0.246
	MSPE	0.930	0.911	0.913	0.930	0.119	0.121	0.121	0.119
	CURE Deviation (%)	7	2	2	6	1	1	1	1

Validation Factor (VF)	1.058	1.036	1.037	1.059	0.996	1.006	0.998	0.997

Notes: MPB: Mean Prediction Bias, MAD: Mean Absolute Deviation, MSPE: Mean Squared Prediction Error

In the case of signalized intersections, the predictive performance of the NB-2 and NB-P models, based on the measures in **table 6,** was better than the NB-1 model for all crash severity levels. Similarly, the NB-2 and NB-P models also outperformed the GP model for total crashes, the PDO crashes, and injury & fatal crashes. Of particular interest was the percentage CURE Deviation, the values of which were very high for all crash severity levels on the signalized intersections in case of the NB-1 and the GP models which clearly shows poor prediction performance. For the unsignalized intersections, the difference among the predictive performance measures across the developed SPFs was extremely small and somewhat inconsequential. However, a close observation shows that the performance of the NB-1, NB-P, and GP models was almost similar and slightly better than the NB-2 model. This finding is in a line with the results of the AIC and BIC (**Table 4**) and the LRT (**Table 5**). To put things into perspective, the GP regression was the best among all four models for each severity level while the NB-P and NB-1 both performed closely and relatively better than the NB-2 model when only the NB models were considered.

5. Discussion

The current study investigated the application of the NB-1, NB-2, NB-P, and a GP model in the development of the SPFs with an aim to find a statistical model capable of improved estimation accuracy. For this purpose, a number of goodness of fit (likelihood-based) and predictive performance measures were calculated and compared. Because several variables were used in the development of the SPFs, a few of them were found to have a significant relationship with the crash occurrence on intersections.

5.1 Predictor variables of crashes on urban intersections

A positive association between the crash frequency and traffic volume of major and minor approaches of intersections was found for almost every severity level and every intersection type considered. This results was in accordance with our expectations. When the number of vehicles entering and/or leaving the intersection increases, it induces new turning and crossing maneuvers, that results in an increased risk of new conflicts, and those additional conflicts, in some cases, are translated into actual crashes. Similar results have been reported by many studies (Wang et al., 2019; Barbosa et al., 2014; Ferreira and Couto, 2013; Vieira Gomes et al., 2012; Miaou and Lord, 2003; Greibe, 2003). It is important to note that, of all the models developed across all severity levels for both intersection types, only the traffic flow of a minor approach of unsignalized intersections in the models for injury & fatal crash was not significant. Since majority of unsignalized intersections were located on local streets where the traffic volume of minor approaches and the corresponding speed limits were relatively very low, it is possible that those factors might have contributed to reduced number of fatal and injury crashes and hence resulted in the insignificance of traffic volume of the minor approaches in models for fatal and injury crashes.

The presence of crosswalk on the minor approaches although significant gave somewhat mixed results for signalized and unsignalized intersections. In the case of signalized intersections, the crosswalk on the minor approaches had a significant positive association with crash frequency only when it was present on both approaches. The crosswalk, however, was not significant when existed on one of the minor approaches. Further, the estimated coefficients were often more than the double for intersections with crosswalks on both minor approaches than intersections with crosswalks on only one approach, although not significant in the later case. A possible explanation could be that, at signalized intersections with crosswalks on both minor approaches, an existing and/or entering or turning traffic will have two possible vehicle-pedestrian interactions and thus the chances of involvement in crashes will be greater, while intersection with a crosswalk on only one minor approach will have one possible vehicle-pedestrian interaction and hence lower risk of a crash. In the case of unsignalized intersection, a crosswalk on a minor approach was significant across all developed models when it was present on only one of the minor approaches. As we know, two crosswalks on the minor approaches were only present on four-legged intersections but in unsignalized category, majority of intersections were three-legged which could accommodate only one crosswalk on its minor approach at a time. Thus, the majority of three legged intersections, and the consequent presence of only one crosswalk on minor approach possibly contributed significantly to crashes on unsignalized intersections.

The intersection skewness was statistically significant in the models for total crashes, and injury & fatal crashes in case of signalized intersections. The association found indicates that more crashes were expected on the intersections with higher skewness level than those with no or lower skewness level. For the recollection of reader, an intersection angle of less than or equal to 75 degrees was higher skewness level and an intersection angle of greater than 75 degrees was lower skewness or no skewness. Nightingale et al. (2017) and Harkey (2013) have reported similar results when studying the influence of the skew angle on the intersection crash frequency. However, Nightingale et al. (2017) studied the rural intersection. The significance of skewness in the SPFs for signalized intersection could be attributed to the fact that drivers tend to have greater perception of safety in the case of signalized intersections than the unsignalized intersections, which reduces an amount decision-making necessary for safe driving. When such drivers encounter a skewed intersection, this potentially lead to confusion, which when reinforced by other undesirable characteristics (obstructed views, distorted sight distances, large intersection area, large turns, etc.) of skewed intersection may result in a crash.

The number of approaches of the intersection was found to influence the expected crash frequency at unsignalized intersections only. This was particularly true in the case of PDO crashes and total crashes. The intersections with four or more legs were expected to experience more crashes than the intersections with three legs. This was expected outcome. An increase in the number of legs/approaches increases the intersection complexity and it invites additional traffic, which could be related to an increased risk of involvement in a crash.

Another significant predictor of crashes at un-signalized intersections was the number of through lanes of the minor approach. The association between the number of through lanes and the expected crash frequency was positive, which means more crashes with an increased number of through lanes. Abdel-Aty and Nawathe (2006) found similar results for urban intersection but their study was focused on signalized intersections. Zhao et al. (2018) and Kamrani et al. (2017) also reported a significant positive association between crash frequency and the number of through lanes for intersections. The number of through lanes on a minor approach also indirectly informs about the size of an intersection and, thus, correspondingly higher traffic volumes. An intersection with a high number of through lanes on a minor approach could have a higher expected crash frequency because of its large size that carries more traffic. This result can be generalized only to four-legged unsignalized intersections since through lanes were only reported for such facility type.

We also found some unexpected results, especially the insignificance of the exclusive left and right turn lanes in the developed models. It was rather opposite to the results of some studies (Al-Kaisy and Roefaro, 2012; Abdel-Aty and Haleem, 2011; Zhou et al., 2010). The reason may be that the number of intersections with exclusive left and right turn lanes in the study data was not enough to be significant in the final models. The influence of the intersection size on crash frequency was also insignificant as a predictor variable. This might be because other variables, like, the number of through lanes on a minor approaches and the number of legs/approaches of the intersection, could have acted as proxies for intersection size in the modelling process.

5.2 The appropriate model(s) for crash estimation on urban intersections

In case of signalized intersections, the comparison of likelihood-based and predictive ability-based measures both revealed that the NB-P and NB-2 models performed better than the NB-1 and GP models. Similarly when compared for the un-signalized intersections, the GP model was a winner based on the goodness of fit (likelihood-based measures), however, the performance of the GP model for predictive ability-based measures was only marginally better than the NB-1 and NB-P models. When the comparison was made among the NB models only for unsignalized intersections, the NB-P and NB-1 performed better than the NB-2. Generally, in situations where only the NB models were considered, the flexible variance structure allowed the NB-P model to outperform the traditional forms, either the NB-1 or NB. Another observation was that for one type of facility (un-signalized intersections), the better performing model was the NB-1, while for the other type of facility (signalized intersection), the better performing model the NB-2 when only the traditional NB models were compared. This finding suggests that it is necessary to check for an appropriate model form in advance.

The use of several functional forms of the NB regression and the equally powerful but a relative less used GP model in our study revealed that the accurate estimation of crash frequency is subjected to the selection of the appropriate functional form and model type. The flexible variance structure of the NB model has the ability to improve the estimation accuracy. Further, the study results showed that it is possible that a model functional form appropriate for one sub-type of the same infrastructure might not be appropriate for another sub-type of that infrastructure.

6. Conclusions

In this study, we developed multiple SPFs by crash severity for urban intersections using the NB-1, NB-2, NB-P and GP regression in an attempt to obtain a model with a higher estimation accuracy. The data was obtained for the intersections of Antwerp, Belgium. Only those intersections were included in modeling for which a sufficient good quality data was available. Major and minor approach AADT and several other variables related to road infrastructure and geometry were used as the explanatory variables. Traffic volume was a significant predictor of crash frequency for almost all developed models and all crash severity levels. Other significant variables include the presence of a crosswalk on the minor approach and the intersection skewness in the case of signalized intersections. For unsignalized intersection, the presence of a crosswalk on the minor approach, the number of through lanes of the minor approach, and the number of legs were significant.

For model comparison, two sets of measures were computed. The likelihood-based measures including the LRT, the AIC and BIC were used for the checking goodness of fit of the models while the predictive ability-based measures were used for the predictive performance and validation of the models. The likelihood-based measures showed that the NB-P and NB-2 models performed better than the NB-1 and GP models in case of the signalized intersections for all crash severity levels. For unsignalized intersections, however, the GP model was relatively better than the NB models. A comparison among the NB models showed that the NB-P and NB-1 outperformed the NB-2. The predictive ability-based measures also confirmed similar results by indicating an improvement in prediction accuracy in case of the NB-P model and the GP model for signalized and unsignalized intersections, respectively.

The findings of this study showed that all functional forms of the NB model and the GP model were promising in the estimation of the SPFs for intersections. The developed models irrespective of the functional form or type showed similar results for the influence of the explanatory variables on crash occurrence. Further, it was shown that the use of the flexible variance structure of the NB-P model and/or an entirely different GP model could bring an improvement in the estimation accuracy as indicated by the comparison of the goodness of fit and later verification by the predictive performance measures applied to the validation dataset.

Finally, it is hoped that the outcome of this study add to the knowledge of the SPF estimation with regard to the selection of the functional form and improvement in the accuracy and reliability of the crash estimates**.** Nonetheless, a future research efforts can focus on investigating the applications of the NB-P model to several other facility types or using the NB-P model in conjunction with other techniques, for instance, exploring the functional forms of the GP model of which a traditional form called GP-1 has already been used in this study.

Acknowledgements

The authors would like to thank the Police of Antwerp for providing the crash data, the Lantis, Antwerp-based mobility management company, for providing the necessary traffic data and the Flemish government for the road infrastructure data used in this research.

References

AASHTO, 2014. Highway Safety Manual (HSM), First Edition, 2014 Supplement, American Association of State and Highway Transportation Officials, Washington DC.

AASHTO, 2010. Highway Safety Manual (HSM), First Edition, American Association of State and Highway Transportation Officials, Washington DC.

Abdel-Aty, M., Haleem, K., 2011. Analyzing angle crashes at unsignalized intersections using machine learning techniques. Accid. Anal. Prev. 43 1 , 461–470. doi:10.1016/j.aap.2010.10.002

Abdel-Aty, M., Nawathe, P., 2006. A Novel Approach for Signalized Intersection Crash Classification and Prediction. Adv. Transp. Stud. 9.

Abdel-Aty, M., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accid. Anal. Prev. 32 5 , 633–642. doi:10.1016/S0001-4575(99)00094-9

Abdel-Aty, M.A., Lee, J., Eluru, N., Cai, Q., Al Amili, S., Alarifi, S., 2016. Enhancing and Generalizing the Two-Level Screening Approach Incorporating the Highway Safety Manual (HSM) Methods, Phase 2 (No. BDV-24-977-06). Florida Department of Transportation.

Al-Kaisy, A., Roefaro, S., 2012. Channelized right-turn lanes at signalized intersections: the U.S. experience 13.

Barbosa, H., Cunto, F., Bezerra, B., Nodari, C., Jacques, M.A., 2014. Safety performance models for urban intersections in Brazil. Accid. Anal. Prev. 70, 258–266. doi:10.1016/j.aap.2014.04.008

Cafiso, S., Di Silvestro, G., Di Guardo, G., 2012. Application of highway safety manual to Italian divided multilane highways. Procedia - Soc. Behav. Sci., SIIV-5th International Congress - Sustainability of Road Infrastructures 2012 53, 910–919. doi:10.1016/j.sbspro.2012.09.940

Cameron, A.C., Trivedi, P.K., 1986. Econometric models based on count data: Comparisons and applications of some estimators and tests. J. Appl. Econom. 1 1 , 29–53.

Chang, G.-L., Xiang, H., 2003. The Relationship between Congestion Levels and Accidents. Maryland State Highway Administration.

Dixon, K., Monsere, C., Avelar, R., Barnett, J., Escobar, P., Kothuri, S., Wang, Y., 2015. Improved safety performance functions for signalized intersections (No. FHWA-OR-RD-16-03). Oregon Department of Transportation.

Dong, C., Richards, S.H., Clarke, D.B., Zhou, X., Ma, Z., 2014. Examining signalized intersection crash frequency using multivariate zero-inflated Poisson regression. Saf. Sci. 70, 63–69. doi:10.1016/j.ssci.2014.05.006

Elvik, R., Sagberg, F., Langeland, P.A., 2019. An analysis of factors influencing accidents on road bridges in Norway. Accid. Anal. Prev. 129, 1–6. doi:10.1016/j.aap.2019.05.002

Famoye, F., Singh, K.P., Wulu, J.T., 2004. On the generalized Poisson regression model with an application to accident data. J. Data Sci. 2, 287–295.

Ferreira, S., Couto, A., 2013. Traffic flow-accidents relationship for urban intersections on the basis of the translog function. Saf. Sci. 60, 115–122. doi:10.1016/j.ssci.2013.07.007

Gates, T., Savolainen, P., Avelar, R., Geedipally, S., Lord, D., Anthony, I., Stapleton, S., 2018. Safety performance functions for rural road segments and rural intersections in Michigan | Blurbs New | Blurbs | Main (No. SPR-1645). Michigan Department of Transportation.

Geedipally, S.R., Lord, D., 2010. Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson–gamma models. Accid. Anal. Prev. 42 4 , 1273–1282. doi:10.1016/j.aap.2010.02.004

Giuffrè, O., Granà, A., Giuffrè, T., Marino, R., Marino, S., 2014. Estimating the safety performance function for urban unsignalized four-legged one-way intersections in Palermo, Italy. Arch. Civ. Eng. 60 1 , 41–54. doi:10.2478/ace-2014-0002

Greene, W., 2008. Functional forms for the negative binomial model for count data. Econ. Lett. 99 3 , 585–590. doi:10.1016/j.econlet.2007.10.015

Greibe, P., 2003. Accident prediction models for urban roads. Accid. Anal. Prev. 35 2 , 273–285. doi:10.1016/S0001-4575(02)00005-2

Haleem, K., Abdel-Aty, M., 2010. Examining traffic crash injury severity at unsignalized intersections. J. Safety Res. 41 4 , 347–357. doi:10.1016/j.jsr.2010.04.006

Hardin, J.W., Hilbe, J.M., 2018. Generalized Linear Models and Extensions, Second Edition. Stata Press.

Harkey, D.L., 2013. Effect of Intersection Angle on the Safety of Intersections. North Carolina State University.

Hauer, E., 2015. The Art of Regression Modeling in Road Safety. Springer International Publishing. doi:10.1007/978-3-319-12529-9

Hauer, E., Bamfo, J., 1997. Two tools for finding what function links the dependent variable to the explanatory variables., in: ICTCT 1997 Conference. Lund, p. 18.

Hilbe, J.M., 2011. Negative Binomial Regression, Second. ed. Cambridge university press, New York. doi:10.1017/CBO9780511973420

Ismail, N., Jemain, A.A., 2007. Handling overdispersion with negative binomial and generalized Poisson regression models. Casualty Actuar. Soc. Forum Citeseer.

Ismail, N., Zamani, H., 2013. Estimation of claim count data using negative binomial, generalized Poisson, zero-inflated negative binomial and zero-inflated generalized Poisson regression models. Casualty Actuar. Soc. E-Forum 41 20 , 1–28.

Janstrup, K.H., 2016. Statistical modelling of the frequency and severity of road accidents. DTU Transport.

Joe, H., Zhu, R., 2005. Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. Biom. J. Biom. Z. 47 2 , 219–229. doi:10.1002/bimj.200410102

Joshua, S.C., Garber, N.J., 1990. Estimating truck accident rate and involvements using linear and Poisson regression models. Transp. Plan. Technol. 15 1 , 41–58. doi:10.1080/03081069008717439

Kamrani, M., Wali, B., Khattak, A.J., 2017. Can data generated by connected vehicles enhance safety?: Proactive approach to intersection safety management. Transp. Res. Rec. 2659 1 , 80–90. doi:10.3141/2659-09

Liu, C., Sharma, A., 2018. Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity. Anal. Methods Accid. Res. 17, 14–31. doi:10.1016/j.amar.2018.02.001

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transp. Res. Part Policy Pract. 44 5 , 291–305. doi:10.1016/j.tra.2010.02.001

Lord, D., Park, B.-J., 2015. Appendix D: Negative Binomial Regression Models and Estimation Methods.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. Anal. Methods Accid. Res. 1 0 .

Melliana, A., Setyorini, Y., Eko, H., Rosi, S., Purhadi, 2013. The comparison of generalized Poisson regression and negative binomial reression methods in overcoming overdispersion. URL /paper/The-Comparison-Of-Generalized-Poisson-Regression-In-Melliana-Setyorini/9e679999bb7585e9965b271dc5afc462b8572114 (accessed 8.21.20).

Miaou, S.-P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accid. Anal. Prev. 26 4 , 471–482. doi:10.1016/0001-4575(94)90038-8

Miaou, S.-P., Lord, D., 2003. Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes Versus Empirical Bayes Methods: Transp. Res. Rec. doi:10.3141/1840-04

Miaou, S.-P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accid. Anal. Prev. 25 6 , 689–709. doi:10.1016/0001-4575(93)90034-T

Nambuusi, B.B., Brijs, T., Hermans, E., 2008. A review of accident prediction models for road intersections. UHasselt.

Nesheli, M.M., Che Puan, O., Roshandeh, A.M., 2009. Evaluation of effect of traffic signal coordination system on congestion. WSEAS Trans. Adv. Eng. Educ. 6 7 , 203–212.

Nightingale, E., Parvin, N., Seiberlich, C., Savolainen, P.T., Pawlovich, M., 2017. Investigation of skew angle and other factors influencing crash frequency at high-speed rural intersections. Transp. Res. Rec. 2636 1 , 9–14. doi:10.3141/2636-02

Oh, J., Kim, E., Kim, M., Choo, S., 2010. Development of conflict techniques for left-turn and cross-traffic at protected left-turn signalized intersections. Saf. Sci. 48 4 , 460–468. doi:10.1016/j.ssci.2009.12.011

Oh, J., Lyon, C., Washington, S., Persaud, B., Bared, J., 2003. Validation of FHWA Crash Models for Rural Intersections: Lessons Learned. Transp. Res. Rec. 1840 1 , 41–49. doi:10.3141/1840-05

Okamoto, H., Koshi, M., 1989. A method to cope with the random errors of observed accident rates in regression analysis. Accid. Anal. Prev. 21 4 , 317–332. doi:10.1016/0001-4575(89)90023-7

Park, B.-J., 2010. Application of finite mixture models for vehicle crash data analysis. Texas A&M University.

Park, B.-J., Lord, D., 2009. Application of finite mixture models for vehicle crash data analysis. Accid. Anal. Prev. 41 4 , 683–691. doi:10.1016/j.aap.2009.03.007

Persaud, B., Saleem, T., Faisal, S., Lyon, C., Chen, Y., Sabbaghi, A., 2012. Adoption of Highway Safety Manual Predictive Technologies for Canadian Highways, in: 2012 Conference of the Transportation Association of Canada. Fredericton, New Brunswick.

Roshandeh, A.M., Agbelie, B.R.D.K., Lee, Y., 2016. Statistical modeling of total crash frequency at highway intersections. J. Traffic Transp. Eng. Engl. Ed. 3 2 , 166–171. doi:10.1016/j.jtte.2016.03.003

Roshandeh, A.M., Levinson, H.S., Li, Z., Patel, H., Zhou, B., 2014. New methodology for intersection signal timing optimization to simultaneously minimize vehicle and pedestrian delays. J. Transp. Eng. 140 5 , 04014009. doi:10.1061/(ASCE)TE.1943-5436.0000658

Srinivasan, R., Carter, D., 2011. Development of Safety Performance Functions for North Carolina (No. FHWA/NC/2010-09). North Carolina Department of Transportation.

Tay, R., 2015. A random parameters probit model of urban and rural intersection crashes. Accid. Anal. Prev. 84, 38–40. doi:10.1016/j.aap.2015.07.013

Vieira Gomes, S., Geedipally, S.R., Lord, D., 2012. Estimating the safety performance of urban intersections in Lisbon, Portugal. Saf. Sci. 50 9 , 1732–1739. doi:10.1016/j.ssci.2012.03.022

Wang, K., Ivan, J.N., Ravishanker, N., Jackson, E., 2017. Multivariate poisson lognormal modeling of crashes by type and severity on rural two lane highways. Accid. Anal. Prev. 99, 6–19. doi:10.1016/j.aap.2016.11.006

Wang, K., Zhao, S., Jackson, E., 2019. Functional forms of the negative binomial models in safety performance functions for rural two-lane intersections. Accid. Anal. Prev. 124, 193–201. doi:10.1016/j.aap.2019.01.015

Wang, X., Abdel-Aty, M., Nevarez, A., Santos, J.B., 2008. Investigation of safety influence area for four-legged signalized intersections: Nationwide survey and empirical inquiry. Transp. Res. Rec. doi:10.3141/2083-10

Washington, S., Karlaftis, M.G., Mannering, F., Anastasopoulos, P.Ch., 2010. Statistical and Econometric Methods for Transportation Data Analysis, Second. ed. Chapman and Hall/CRC.

Winkelmann, R., 2008. Econometric Analysis of Count Data, 5th ed. Springer-Verlag, Berlin Heidelberg. doi:10.1007/978-3-540-78389-3

Wong, S.C., Sze, N.N., Li, Y.C., 2007. Contributory factors to traffic crashes at signalized intersections in Hong Kong. Accid. Anal. Prev. 39 6 , 1107–1113. doi:10.1016/j.aap.2007.02.009

Wu, Y., Abdel-Aty, M., Cai, Q., Lee, J., Park, J., 2018. Developing an algorithm to assess the rear-end collision risk under fog conditions using real-time data. Transp. Res. Part C Emerg. Technol. 87, 11–25. doi:10.1016/j.trc.2017.12.012

Wu, Z., Sharma, A., Mannering, F.L., Wang, S., 2013. Safety impacts of signal-warning flashers and speed control at high-speed signalized intersections. Accid. Anal. Prev. 54, 90–98. doi:10.1016/j.aap.2013.01.016

Yang, Z., Hardin, J.W., Addy, C.L., 2009. A score test for overdispersion in Poisson regression based on the generalized Poisson-2 model. J. Stat. Plan. Inference 139 4 , 1514–1521. doi:10.1016/j.jspi.2008.08.018

Zhao, M., Liu, C., Li, W., Sharma, A., 2018. Multivariate Poisson-lognormal model for analysis of crashes on urban signalized intersections approach. J. Transp. Saf. Secur. 10 3 , 251–265. doi:10.1080/19439962.2017.1323059

Zhou, H., Ivan, J.N., Sadek, A.W., Ravishanker, N., 2010. Safety effects of exclusive left-turn lanes at unsignalized intersections and driveways. J. Transp. Saf. Secur. 2 3 , 221–238. doi:10.1080/19439962.2010.502613